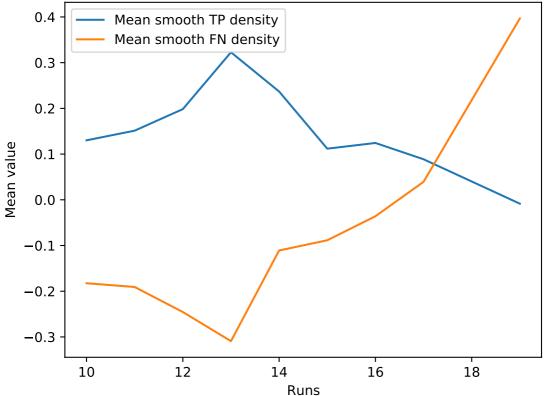
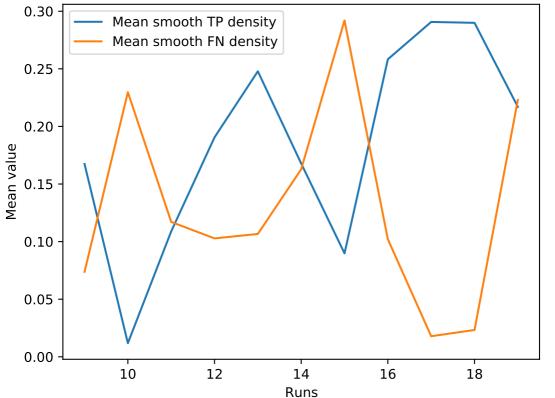


Mean value density difference in positive instances no prediction influence

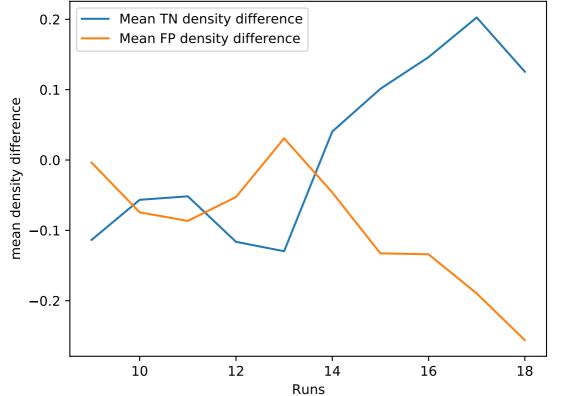


Mean value density difference in positive instances self fulfilling approach -0.20Mean smooth TP density Mean smooth FN density -0.25-0.30Mean value -0.35-0.40-0.45-0.50-0.5514.0 15.0 16.0 16.5 18.0 14.5 15.5 17.0 17.5 Runs

Mean value density difference in positive instances self defeating approach



Mean value density difference in negative instances no prediction influence

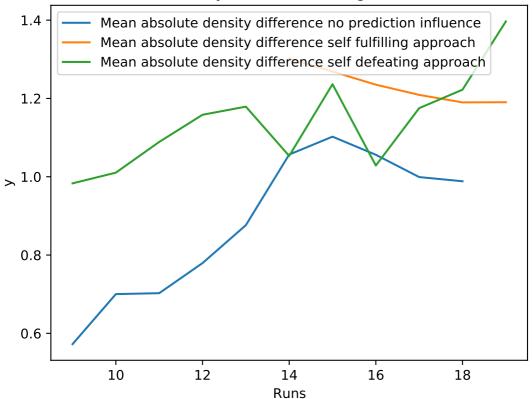


Mean value density difference in negative instances self fulfilling approach Mean TN density difference -0.1Mean FP density difference -0.2mean density difference -0.3-0.4-0.5-0.614 15 16 17 18 19 Runs

Mean value density difference in negative instances self defeating approach 0.35 Mean TN density difference Mean FP density difference 0.30 0.25 mean density difference 0.20 0.15 0.10 0.05 0.00 -0.0510 16 18 12 14

Runs

## Absolute density difference in negative instances



## Absolute density difference in positive instances

